

ECON832 Final: Mini Paper

Choice Prediction Competition 2018 modelling with deep feedforward neural networks

Daniel Sánchez Pazmiño

April 2024

1 Overview

The data comes from the CPC18 (Plonsky, Erev, and Ert 2017), which consists of experiments involving decision makers choosing between two lotteries. I focus on data exclusive to CPC18 (Experiment 1), which added attention features to the CPC15 data (Erev, Ert, and Plonsky 2017). This experiment involved sixty choice problems for 240 participants. I use the calibration data set to train my deep feedforward neural network and the **individual track data** (Track II) to test the model. All replicating code and files are available on the project's GitHub repository.

2 Methodology

2.1 Data preparation

The calibration dataset was filtered to only include Experiment 1 data (sets 5 and 6). The outcome variable was the binary indicator of the choice of B over A. Since the variable was already dichotomous, no transformation was needed. I include demographic variables in the baseline models: location, gender and age. I dichotomized categorical variables with Technion and male as the reference levels.

Risk features involve the number of lottery outcomes, the expected values, the low payoffs, and probabilities to draw from the lotteries and lot distribution shapes. The latter defines the distribution of the lotteries, which can be binomial around the expected value, right skewed or left skewed. They were dichotomized in the same way as the location and gender variables.

Other relevant variables involve whether there is ambiguity in the probabilities or if there is a correlation between the payoffs. Further, it is important to include trial numbers, potential and obtained payoffs obtained, and forgone payoffs. Some of these variables are categorical, but they are already dichotomized, so I didn't transform them.

Regarding attention features, I include the number of time block, reaction time, feedback binary outcome, the onscreen button and the trial number within a game.

The processing of the calibration dataset was done using `TidierData.jl` and `DataFrames.jl`. I performed a test-train split with 80% of the data used for training and 20% for testing. I further tested with the individual track data Plonsky, Erev, and Ert (n.d.).

2.2 Feature engineering

As mentioned, all categorical variables for both the risk and attention features were dichotomized. This is equivalent to **one-hot encoding**. The continuous variables were **standardized** to have $\mu = 0$ and $\sigma = 1$. The data was processed from its rectangular form (every row per decision) to a format where each row represented an attribute, and every column represented a decision.

2.3 Model configuration

The model is a deep feedforward network neural network (DFNN) with `Flux.jl`. I use a ReLU activation function for the input and hidden layers and a sigmoid activation function for the output layer. I use MSE as the loss function and the gradient descent optimizer to minimize the loss function. The learning rate was set to 0.7, and I train the model for 100 epochs using a `for` loop. I use a confusion matrix and an accuracy rate (percent correctly predicted) to evaluate the model's fit, using a 0.5 probability threshold to determine predicted classes.

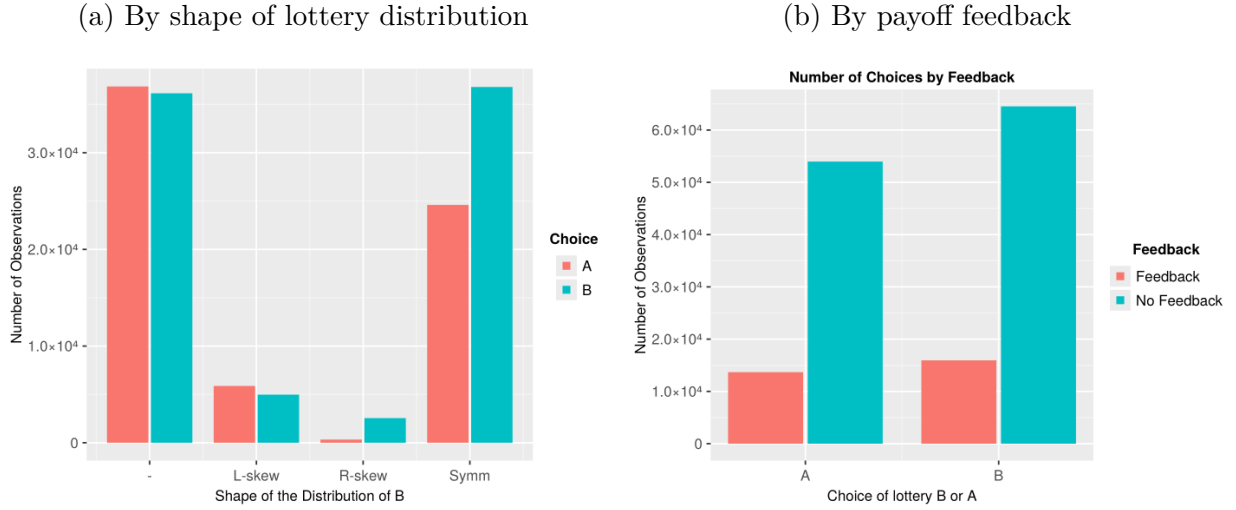
3 Feature analysis

3.1 Risk features

The correlation between the number of outcomes and the choice for the B lottery is 0.09, which is statistically significant at the 1% level. Further, the correlation between the low payoff of the lotteries and the choice for the B lottery is 0.11, which is statistically significant at the 1% level. This might be a result of the risk aversion of the participants, which is a common feature in decision-making and point towards the importance of considering risk-aversion in the model rather than risk neutrality. The correlation between lotteries shows a relationship with the choice of B, with negative correlation of -0.14, which is statistically significant at the 1% level.

In terms of the shape of the distribution, I identify that a symmetric distribution greatly favours the choice of B, as seen in Figure 1a. Further, I also identify that the payoff that the participant received shows a statistically significant correlation with the choice of B. I choose to include all 12 variables which jointly define the shape of the distribution in

Figure 1: Choices of lottery by attention and risk features



the model, as well as the payoff that the participant received plus the potential payoffs of A and B.

3.2 Attention features

I find that the forgone payoff has a statistically significant correlation with the choice of B, with a correlation of -0.016. While reaction time does not show a significant correlation with the choice of B, I include it along time blocks in the interest that its inclusion might point towards the importance of attention, and particularly fatigue, in the decision-making process. Giving feedback to the participant also shows a notable correlation with the choice of B, as shown in Figure 1b. Finally, a button placed toward the left marginally favours choice of A, with a correlation of -0.002 with the choice of B.

4 Results and discussion

Table 1a and Table 1b display confusion matrices for the baseline model with risk features only, for both the training data and the competition data. The accuracy rate for the training data is 75.88%, while for the competition data it is 82.51%. This suggests that the model is not overfitting, as the accuracy rate is similar for both datasets. The loss value (MSE) for the training data is 0.18, while for the competition data it is 0.14 too.

Table 1: Confusion matrices for the baseline DFNN model

(a) Training data			(b) Competition data		
	A	B		A	B
A	28502	25579	A	502	284
B	3007	61412	B	44	1045

The model which considers both risk and attention features has an accuracy rate of 57.12% for the training data and 62.40% for the competition data. The loss function value for the training data is 0.246, while for the competition data it is 0.245. Table 2a and Table 2b display the confusion matrices for the model with attention features.

Table 2: Confusion matrices for the DFNN model with attention features

(a) Train			(b) Competition data		
	A	B		A	B
A	4844	28779	A	120	666
B	1445	35423	B	39	1050

These performance metrics suggest that adding attention features to the feedforward neural network model does not improve the model’s accuracy. Attention features in this context may not be as relevant as the risk features in predicting the choice of B over A, and likely add noise to the model rather than improve its predictive power. The model with risk features only is the best model to predict the choice of B over A in the way that I have defined it.

The deep forward feedward neural network model does not impose any structure on the data, unlike economic models such as BLP, which defines an indirect utility function $u_{ij} = x_{ij}\beta + \xi_{ij}$, where x_{ij} is the vector of observed characteristics of the choice, β is the vector of parameters to be estimated, and ξ_{ij} is the unobserved utility. The neural network model is more flexible and can capture non-linear relationships between the features and the outcome variable, however, it does not understand the underlying economic mechanisms that drive the decision-making process. This is why DFNNs are famously hard to interpret, unlike economic models, which typically provide clear interpretations of final results. For instance, using BLP yields an interpretable ξ parameters which would allow me to understand the importance of each feature in the decision-making process. If I were to observe that the ξ parameter for the number of outcomes is negative, I could infer that the more outcomes a lottery has, the less likely the participant is to choose it. This is purely theoretical given that BLP works with supply-side administrative data, however, it does speak toward the importance of understanding the underlying economic mechanisms that drive the decision-making process.

In reviewing the literature which relates to the CPC18, it was seen that the best performing models were those which used a combination of machine learning and behavioural models

(which take into account theory which we use in economics). CPC15 was won by the *Psychological Forest*, which random forests and behavioural models in conjunction (Plonsky et al. 2017). This model, with a 0.14 MSE, compares to the 0.13 MSE of the neural net models which were used as benchmarks in this paper (See table 10). Thus, incorporating theoretical mechanisms likely improves the model’s predictive power, however, only relying on structural models is still not enough to maximize prediction power.

References

- Erev, Ido, Eyal Ert, and Ori Plonsky. 2017. “Raw Data for CPC2015: A Choice Prediction Competition for Decisions Under Risk, Under Ambiguity, and from Experience.” Data set with codebook. Zenodo. <https://doi.org/10.5281/zenodo.321652>.
- Plonsky, Ori, Ido Erev, and Eyal Ert. 2017. “Calibration Data for Choice Prediction Competition 2018 (CPC18).” Data set with codebook. Zenodo. <https://doi.org/10.5281/zenodo.845873>.
- . n.d. “Competition Set Data Track II (Individual Behavior): Raw Data.” Data set. Zenodo. Accessed April 20, 2024. [CPC18: Choice Prediction Competition 2018](#).
- Plonsky, Ori, Ido Erev, Tamir Hazan, and Moshe Tennenholtz. 2017. “Psychological Forest: Predicting Human Behavior.” In. Vol. 31. 1.